EXTREME VALUE THEORY

Richard L. Smith

STOR Department: Advanced Special Topics Course

Background: Extreme Value Theory is the branch of probability theory and statistics that is concerned with extremes of sequences of random variables and stochastic processes. The subject started in the 1920s with a seminal paper by Fisher and Tippett and an almost-parallel contribution by Fréchet, but was first put on a firm mathematical ground by Gnedenko (1943). The original theory was for independent and identically distributed (IID) random variables but it was quickly extended to other cases such as independent non-identically distributed random variables, stationary (dependent) sequences and continuous-parameter stationary processes. Beginning in the 1950s but more prominently in the 1970s, the theory started to be extended to multivariate random variables and more recently to general stochastic processes including spatial and spatio-temporal processes. Statistical methods for extremes were originally developed in a classic but now outdated book by Gumbel (1958), but were developed much more extensively in the 1980s with the increased availability of computers and automated computational algorithms (e.g. fitting the Generalized Extreme Value distribution by maximum likelihood). These days, the theory is widely applied in many fields but three of the main ones are reliability and strength of materials, mathematical finance and insurance, and environmental (especially, climate) extremes. A side interest of mine is applications to records in athletic events.

The aim of the course is twofold – on the one hand, to cover the mathematical theory (probabilistic and statistical) with some degree of rigor, and on the other, some of the major areas of application. The mathematical theory will cover broadly four categories – IID sequences, stationary sequences, multivariate and spatial – and the applied fields will include each of the reliability, insurance/finance and environmental/climate fields. The last of these has been my own main applied statistics interest in recent years so we will certainly be covering that topic in some detail.

Instructor. Richard L. Smith, Hanes 303, rls@email.unc.edu (email is the preferred method of communication). I have office hours each week from 12:00-2:00 Mondays and Wednesdays. These were established primarily for students in my other course this semester, STOR 590, but I would be happy to meet STOR 834 students at this time as well. Alternatively, please email me for an individual appointment (always encouraged). There will not be an IA or grader for this course.

Format of Course. The course will meet twice a week in Hanes 107, Tuesday and Thursdays 12:30-1:45, first class January 9, last class on Thursday, April 23. There will be no class March 10 or 12 because of Spring Break. As of now, I am not aware of any other breaks in the class schedule but it's possible I may need to make some changes – if so they will be announced well ahead of time.

Class Materials. Since much of the course will be based on notes that I or other people have written at different times, or that I hope to develop during the course, or past lectures of mine, I want to make these available to the class. My intention is to post these materials on the course webpage, at

https://rls.sites.oasis.unc.edu/s834-2020/s834.html

I have also created a sakai page for this course and will post materials there (under the "Resources" tab) if I don't want to post them publicly at the present time. All registered students in the course will automatically be added to the sakai page. If you are not a registered student and would like access, please ask me.

To start off the course, I plan to review the following materials: the lectures I gave at NCSU in 2015

https://rls.sites.oasis.unc.edu/EVT2015.html

and a longer book chapter based on my SEMSTAT notes from 2001 (which covers some of the same material), on my webpage at

https://rls.sites.oasis.unc.edu/postscript/rs/semstatrls.pdf

However, I also want to cover recent work by people such as Anthony Davison, Jonathan Tawn, Jenny Wadsworth and Dan Cooley. Materials will be posted on sakai or the course webpage as appropriate.

In addition, there are numerous books on extreme value theory, which run the gamut from very theoretical books on the probability and stochastic processes side to the subject, to more applied books focusing on statistics. A partial (mostly optional) reading list is given at the end of this syllabus.

Assessment. There will be no homeworks or written exams in this course. The sole assignment will be a short project due at the end of semester. I am flexible about the format of this project but my main expectation is that it will be either (a) a review of literature or some other topic outside what I present in class, or (b) some independent research (which could be data analysis) of your own. I will expect you to present a report on your project in class and to hand in a short (4-6 pages) report about it. If you attend the class with reasonable regularity and complete the project you will get a P grade in the class. H grades will be awarded at my discretion to students who do a superior job on the project.

As an alternative to this form of project, I would like to suggest one other possibility. I only just received a notice of the paper by Engelke and Hitz (2020), which is being presented as a "read paper" to the Royal Statistical Society on February 12. By long tradition of the Royal Statistical Society, discussion contributions are accepted and printed in the journal along with the authors' reply. These are unrefereed and anyone may contribute so long as they are within the 400 word limit (not counting figures and tables). To be considered for publishing, contributions must be submitted by February 26.

So my alternative suggestion for a project is: *Write a 400-words-or-less discussion of this paper.* I will not insist that it actually has to be submitted to the journal, but would encourage anyone who thinks about this possibility to do that if you can complete it in time. The paper is about network models for extremes but the main prerequisite for reading it is the theory of extremes in (low-dimensional) multivariate cases. I certainly intend to cover that topic but it will not be right at the beginning to the semester.

If you do decide to do that (whether or not you submit it the journal – you have to submit it to me to get credit) your grade for the course will be based solely on that and there will be no further requirement for a project (though I may ask you to present your discussion in class).

Outline of topics:

1. The Extreme Value Limit Distributions. The basic theory of extremes of univariate IID random variables as formulated by Fisher and Tippett, Gnedenko, and de Haan. The extreme value

distributions and their domains of attraction; rates of convergence; statistical theory of the Generalized Extreme Value and Generalized Pareto distributions.

2. Extremes in Stochastic Sequences. Leadbetter's D and D' conditions; the extremal index; computations of extremal index in non-trivial situations, including Markov chains. Estimation of extremes under dependent conditions.
3. Multivariate Extremes. Multivariate extreme value distributions and their domains of attraction. Asymptotic dependence and asymptotic independence. Estimation under both parametric and semiparametric families.
4. Spatial Extremes. Max-stable and max-infinitely divisible processes; constructions and characterizations. Special classes such as Brown-Resnick processes. Statistical theory: method of composite likelihood; progress towards exact maximum likelihood; Bayesian approaches.
5. Applications to Reliability and Strength of Materials.
6. Applications to Finance and Insurance.
7. Applications to Climate and Environment.

References (partial list):

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004), *Statistics of Extremes: Theory and Applications*. Wiley, Chichester, England.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*. Springer, New York.

Cramér, H. and Leadbetter, M.R. (1967), *Stationary and Related Stochastic Processes*: *Sample Function Properties and Their Applications.* Wiley, New York.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, New York.

Gumbel, E.J. (1958), *Statistics of Extremes*. Columbia University Press, New York (of historical interest only!)

Haan, L. de and Ferreira, A. (2006), *Extreme Value Theory: An Introduction*. Springer, New York.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Series.* Springer-Verlag, New York.

Resnick, S. (1987), *Extreme Values, Point Processes and Regular Variation*. Springer-Verlag, New York.

Resnick, S. (2007), *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.

R.L. Smith (2003), Statistics of extremes, with applications in environment, insurance and finance. Chapter 1 of *Extreme Values in Finance, Telecommunications and the Environment*, edited by B. Finkenstadt and H. Rootzén, Chapman and Hall/CRC Press, London, pp. 1-78.

Royal Statistical Society Paper:

Sebastian Engelke and Adrien S. Hitz, (2020), Graphical models for extremes. To be presented at a meeting of the Royal Statistical Society on February 12, 2020. Preprint available from:

https://www.rss.org.uk/Images/PDF/publications/2019/Engelke-12-Feb-2020.pdf